



# Defining Data Warehousing

What is it and who needs it?



---

**A Silvon Software, Inc.  
White Paper**

---

# Defining Data Warehousing

## What is it and who needs it?

*The concept of data warehousing began in the late 1980s. The principal was and still is that fundamental business systems are designed to capture as much information as possible, work as quickly as possible, and have very low failure rates. Note that user accessibility is not a top priority. Hence, a business system (general ledger, order entry, inventory control, etc.) collects a veritable treasure of information, but that information is stored in a system that was never designed for access by businesspeople.*

The “data warehouse” is supposed to be a repository for the useful data collected by those business systems. The difference is that the data warehouse is designed for easy user access. This difference solves two constant problems:

- The ability of business decision-makers to gain timely access to corporate information — It is ironic, but not surprising, that the very transaction-based systems on which the business depends most heavily are designed in such a way that IS generally is the only department capable of effective access & retrieval. Most transaction systems are designed with hundreds of obscure table and field names (the design target is volume and speed). Studies have shown that users find it difficult to locate data when there are more than 22 tables in a database. You see the conflict.
- Consistency of results — if everyone uses the same data source, report results should be synchronized. Result inconsistency arises because of the very nature of transaction systems. These on-line systems constantly change content as new data is added. The stories you usually hear are that IS “doesn’t want users accessing the transaction systems” because it affects performance. The truth is that accessing the transaction systems for reporting is as bad for the users as it for IS. Here’s why. Let’s say my boss and I both run a sales report program against the transaction system. Mine runs at 10:00 AM, and shows we are 25% under quota. My boss runs his report at 12:15 PM. We are scheduled to have a staff meeting at 1:00 PM where I will discuss sales trends. What I don’t know is that at 11:00 AM, a sale valued at \$50,000,000 was entered into the system. So my presentation is geared to addressing the shortcomings of sales, while the Sales VP is looking at numbers that show us outperforming quota by 75%.

## Data warehouse implementations

---

The traditional data warehouse is a centrally maintained data storage facility that (in theory) holds all the data any user could want. These systems are inevitably technically driven. Cost is high (\$3,000,000 - \$5,000,000). Implementation time is long (12-18 months). These gigantic enterprisewide undertakings are falling out of favor because of expense, time, and failure rates.

## The data mart option

---

Instead, businesses are opting for data marts. These implementations serve the same function as a data warehouse (ease of user access, data synchronization, etc.), but are aimed at specific business problems, generally within departments or divisions. Costs are much lower (typically well under \$500,000). Implementation time is less than 6 months (assuming the implementer has experience). The data mart can be accessed by a huge range of user tools, from spreadsheets like Excel and 1-2-3 to sophisticated analysis tools.

Recently, data mining has become yet another option. Customers get confused by data mining, so let's unconfuse" this technology. Data warehouse/data mart is a solution for the problems of information analysis by business decision makers. Data mining is a method for determining patterns in your business operations. A data mining implementation accumulates enormous volumes of transaction data. The data mining system then sifts through that data looking for some type of pattern. Data mining is most heavily used by retailers looking for buying patterns in point-of-sale data. At the moment data mining is in its infancy. It is very expensive, with highly specialized user tools.

## Building the data warehouse/data mart

---

The steps for building a data warehouse/data mart are straightforward. Here we will concentrate on the data mart, as that is now 75% of the market.

- First, you must determine a user “driver.” Data marts are rarely successful if they are driven by technology groups.
- Determine what information the users want and need.
- Find out where the source data is located.
- Design the data mart DBMS.
- Extract the required data from the source data files.
- Convert/reformat the source data to an acceptable format for the target data mart.
- Load the data mart.

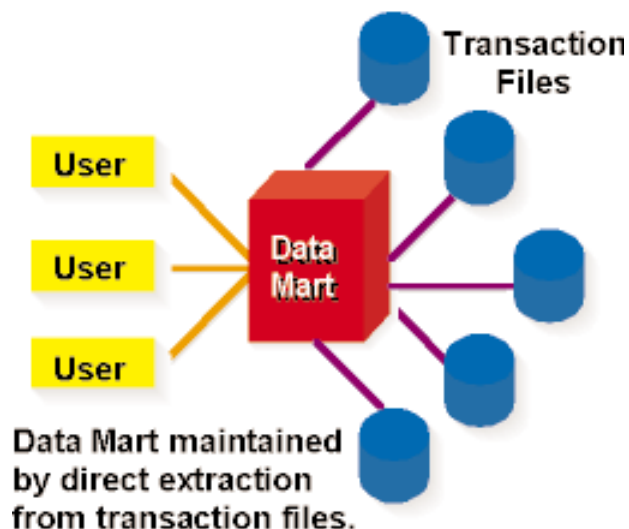
Somewhere along the line, you will probably find a substantial number of errors in your source transaction files. These can be addressed with “data cleansing” tools, if this is an issue.

## Other data warehouse/data mart issues

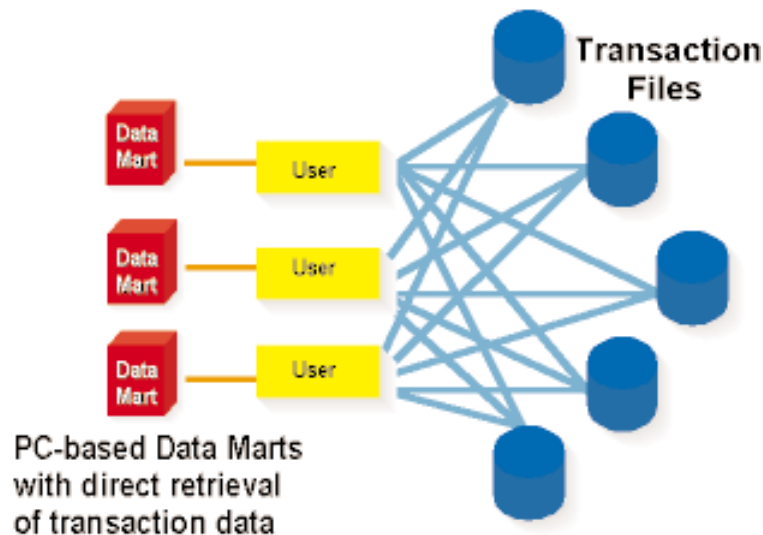
---

Data warehousing is getting to be a crowded market. Vendors are fighting very hard for sales. The hype level is quite high. Be thorough.

- Scalability is a major issue. Once installed, the usage and size of the data mart will often rise geometrically. Storage is cheap, as are new client licenses. The issue that's rarely discussed is loading time. Databases can scale up for storage. Systems can scale up for more users. How about your operational window for updating the data mart — does it scale up? Most customers have, at most, an 8-10 hour operational window for updating the data mart. If the data volumes increase, can the data mart vendor successfully update more data within the same window? If not, the rest of the scalability issues are moot, and you are dead-ended.
- Implementation is another significant issue. Some vendors offer full implementation services for data marts. Others sell you products, and you develop your own expertise or hire consultants. If this is a vertical implementation (manufacturing-specific, food industry-specific, etc.), be sure the implementer has experience in your vertical. Knowing data mart technology is only half the battle. Knowledge of your industry is a big plus toward a successful implementation.
- Product architecture is significant. There are two main approaches being taken by the data mart and OLAP vendors:
  - A server-based approach. The data mart is loaded by extracting data from the transaction system at regular intervals. Users then employ PC client software to analyze the data resident on the data mart server.

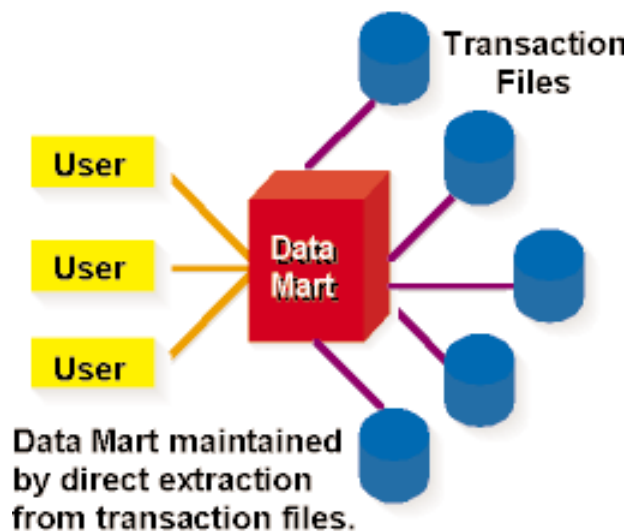


- A client PC-based approach. The PC-based approach puts both the analysis software and the multidimensional data mart on the PC. Users extract the data directly from the transaction system, and load to a PC-based analysis system.

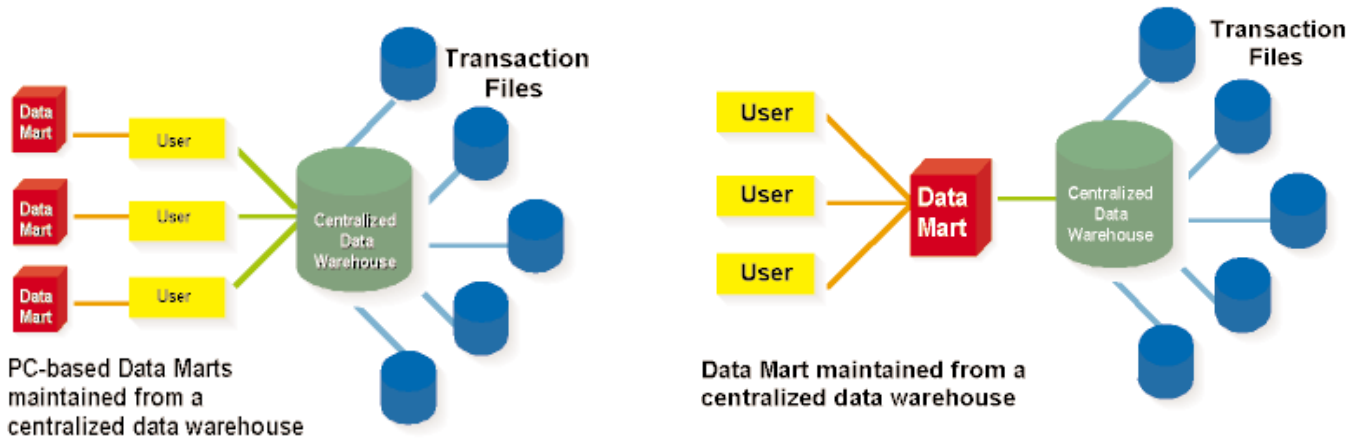


The problem with the PC-based approach is that each user determines when he updates his local database. So again, we could have literally hundreds of different report results. If synchronizing your reports is a major issue, you will be better served with a server-based approach where the data mart is accessed directly by users.

- The data mart implementation itself has two options:
  - Create and maintain the data mart with direct extracts from the transaction systems



- Create an enterprise data warehouse first, then extract from this centralized data warehouse to local data marts.



As you can see in the graphics above, the centrally controlled data warehouse/client approach can also be configured with a single data mart or PC-based data marts based on centralized data warehouse extracts. The same concerns and benefits apply here. With a single data mart, all users are looking at the same data. Individual PC-based data marts offer greater flexibility with the tradeoff in data synchronization

- Is everybody looking at a data representation of the same moment-in-time?
- Are the updates to the centralized warehouse synchronized so that all related data is updated at the same time?

This is often a political issue. A centralized warehouse must be implemented by IS, while the data mart can be implemented by business users. The data mart updates will be dependent upon the centralized warehouse updates. Changes to the centralized data warehouse will need to be propagated to all the data marts downstream. Centralized data warehouses are inherently a more complicated, more technology-driven solution.

## Benefits of data warehousing

Studies have consistently shown a huge return on investment from data warehouse/data mart implementations. The figures generally range from 300% of data warehouses to 650% for data marts.

Business is changing faster than ever, and competition is more brutal than ever. Business decisions require consistent accurate information. The old approach of asking IS to design a report just won't cut it anymore. With hardware prices falling, there has never been a better time to create some autonomy for the business user. Users need direct access to corporate information to get the facts to make business decisions, and data marts are the most cost-effective way to provide that access.

## Data warehousing definitions

---

- Data warehouse — a repository of business information used for analysis, separate from the “active” business system files.
- Data mart — business-focused data warehouse. Usually focuses on a particular business issue.
- Data mining — analysis techniques designed to detect patterns in the data. Could be applied to data mart/data warehouse contents.
- OLAP (On-Line Analytical Processing) — Tools and applications that support interactive analysis of multidimensional data.

## For More Information

---

Silvon Software, Inc.  
900 Oakmont Lane  
Suite 400  
Westmont, IL 60559  
(630) 655-3313  
info@silvon.com

Silvon Software Ltd.  
Ocean House, The Ring  
Bracknell, Berkshire  
RG12 1AX England  
(44) 1344 354900  
info@silvon.com

Silvon Software Pty Ltd.  
138 Harris Street  
Pymont NSW 2009  
Australia  
(61) 2-9552-1216  
info@silvon.com